Developmental Biology **244,** 257–266 (2002)
doi:10.1006/dbio.2002.0603, available online at http://www.idealibrary.com on IDEAL®

Requires an Evolutionarily Conserved Smad Binding Site

Ching-Ling Lien,*,[1] John McAnally,* James A. Richardson,† and Eric N. Olson*,[2]

Department of Molecular Biology, and †Department of Pathology, University of Texas Southwestern Medical Center at Dallas, 6000 Harry Hines Boulevard, Dallas, Texas 75390

Heart formation in vertebrates and fruit flies requires signaling by bone morphogenetic proteins (BMPs) to cardiogenic mesodermal precursor cells. The vertebrate homeobox gene *Nkx2–5* and its *Drosophila* ortholog, *tinman*, are the earliest known markers for the cardiac lineage. Transcriptional activation of *tinman* expression in the cardiac lineage is dependent on a mesoderm-specific enhancer that binds Smad proteins, which activate transcription in response to BMP signaling, and Tinman, which maintains its own expression through an autoregulatory loop. Here, we show that an evolutionarily conserved, cardiac-specific enhancer of the mouse *Nkx2–5* gene contains multiple Smad binding sites, as well as a binding site for Nkx2–5. A single Smad site is required for enhancer activity at early and late stages of heart development *in vivo*, whereas the Nkx2–5 site is not required for enhancer activity. These findings demonstrate that *Nkx2–5*, like *tinman*, is a direct target for transcriptional activation by Smad proteins; however, the independence of this *Nkx2–5* enhancer of Nkx2–5 binding suggests a fundamental difference in the transcriptional circuitry for activation of *Nkx2–5* and *tinman* expression during cardiogenesis in vertebrates and fruit flies. © 2002 Elsevier Science (USA)

INTRODUCTION

Recent studies indicate that heart formation is controlled by an evolutionarily ancient developmental program, many aspects of which are conserved between vertebrates and insects (reviewed in Bodmer, 1995; Fishman and Olson, 1997). In vertebrates, heart formation begins soon after gastrulation, when cells within a bilaterally symmetric region of the anterior lateral plate mesoderm, known as the cardiac crescent, adopt a cardiac fate in response to signaling from the adjacent endoderm (reviewed in Nascone and Mercola, 1996). Studies performed in amphibian and chick embryos have demonstrated that bone morphogenetic proteins (BMPs) mediate cardiogenic induction by the endoderm (Schultheiss *et al.*, 1995, 1997). Similarly, formation of the heart-like organ in *Drosophila*, the dorsal vessel, is dependent on the BMP-related factor, Dpp, which is secreted by the ectoderm and induces underlying dorsal mesodermal cells to become cardiogenic (Frasch, 1995; Xu *et al.*, 1998).

BMP belongs to the transforming growth factor β (TGFβ) superfamily. Signaling by this superfamily of growth factors is mediated by effectors, called Smad proteins. Three classes of Smads have been defined: the receptor-regulated Smads (R-Smads), the co-Smad (Smad4), and the inhibitory Smads (I-Smads). BMP signaling results in phosphorylation of cytoplasmic R-Smads 1, 5, or 8, which then interact with Smad4, followed by translocation of the heteromeric Smad complex to the nucleus, where it associates with other transcription factors to activate specific BMP-responsive genes (reviewed in Attisano and Wrana, 2000; Massague and Chen, 2000). Smad proteins bind DNA relatively weakly alone, but are recruited to specific target genes via their interactions with distinct transcriptional cofactors. The ability of BMP signaling to induce specific mesodermal cells to commit to a cardiac fate requires the interpretation of BMP signals in a cell type-specific manner. Understanding the mechanism that mediates this selective response to

[1] Present address: Department of Cardiology, Children's Hospital, Howard Hughes Medical Institute. 320 Longwood Avenue, Enders Building, Room 1250, Boston, MA 02115.

[2] To whom correspondence should be addressed. Fax: (214) 648-1196. E-mail: eolson@hamon.swmed.edu.

0012-1606/02 $35.00

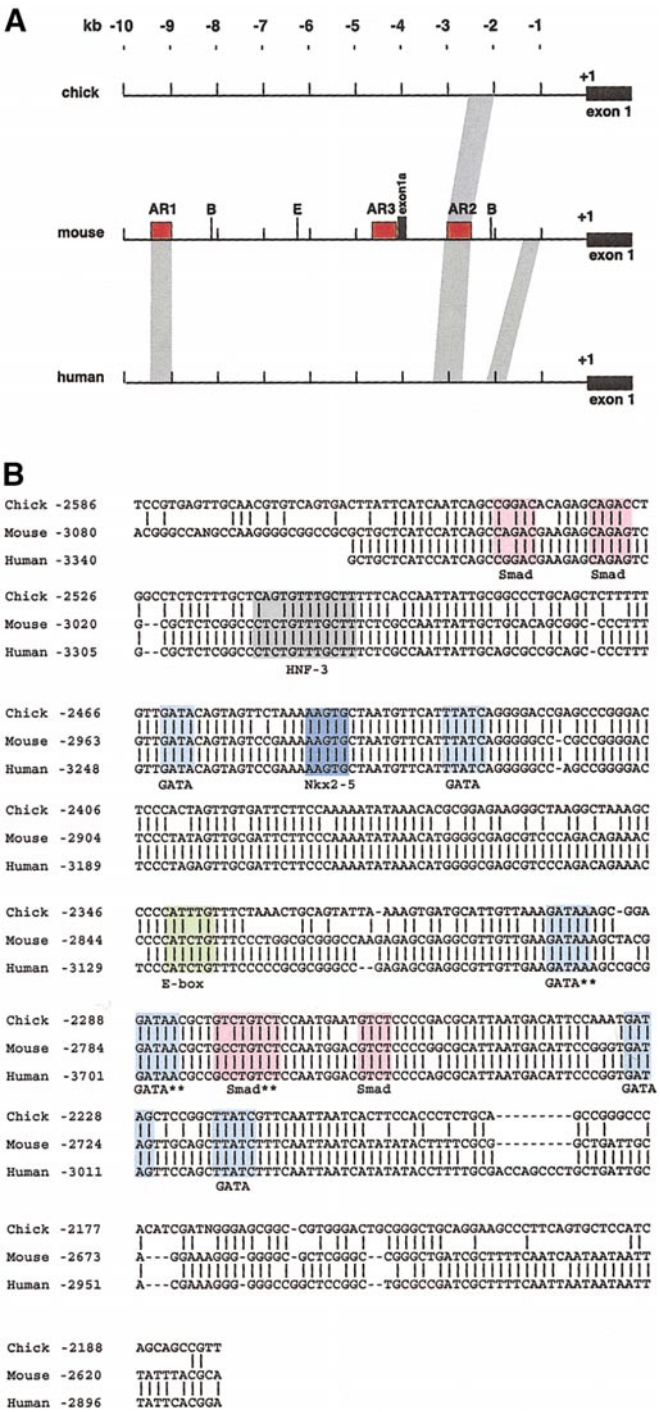

FIG. 1. Cross-species comparison of *Nkx2–5* 5′-flanking sequences (A). Schematic diagram of 5′-flanking regions of chicken, mouse, and human *Nkx2–5* genes. The AR1, AR2, and AR3 enhancers (Schwartz and Olson, 1999) are in red, and regions of homology are indicated in gray. An additional homologous region between mouse and human sequences surrounding −1285 and −2130, respectively, has not been shown to possess transcriptional activity. The transcriptional start site is marked as +1. Exon 1a is an alternative 5′ exon. B, *Bam*HI; E, *Eco*RI. (B). Comparison of the mouse *Nkx2–5* AR2 enhancer with chicken and human sequences. Binding sites for known cardiac transcription factors are marked. The two essential GATA sites and the essential Smad site are marked with asterisks (**).

BMP signaling is likely to provide insight into the molecular basis for cardiac specification, which is poorly understood.

The earliest molecular marker for cardiac development in *Drosophila* is the NK-type homeobox gene *tinman* (Bodmer *et al.*, 1990; Azpiazu and Frasch, 1993; Bodmer, 1993). *tinman*, which is essential for heart formation, is initially expressed throughout the entire mesoderm, before becoming restricted to dorsal mesodermal cells that give rise to the dorsal vessel (Bodmer *et al.*, 1990; Azpiazu and Frasch, 1993; Bodmer, 1993). Activation of *tinman* transcription by Dpp is mediated by binding of the Smad4 ortholog Medea to a dorsal mesoderm-specific enhancer (tin-D) that also binds other activators and repressors, including the Tinman protein itself (Xu *et al.*, 1998). Binding of Tinman to this control region fulfills a positive autoregulatory loop that reinforces and maintains *tinman* expression. However, the Tinman binding sites are not required for Dpp responsiveness of the tin-D enhancer (Xu *et al.*, 1998).

By comparison, the vertebrate ortholog of *tinman*, *Nkx2–5* (Komuro and Izumo, 1993; Lints *et al.*, 1993), is expressed throughout the cardiac crescent concomitant with cardiogenic induction (Harvey, 1996). Attempts to rescue heart formation in *tinman* mutant embryos using vertebrate NK-type homeobox genes have shown that the functions of *Nkx2–5* and *tinman* are at least partially conserved (Park *et al.*, 1998; Ranganayakulu *et al.*, 1998). *Nkx2–5* transcription has been shown to be controlled by a complex series of positive and negative regulatory elements. However, despite embryological data directly implicating BMP signaling in vertebrate cardiogenesis (Schultheiss *et al.*, 1997; Shi *et al.*, 2000), it is unclear whether BMP signaling and its effectors act directly on the *Nkx2–5* gene or whether there are intermediate steps in *Nkx2–5* activation by BMP.

Two cardiac enhancers, referred to as activating region (AR) 1 and AR2, in the 5′-flanking region of the mouse *Nkx2–5* gene (reviewed in Schwartz and Olson, 1999), direct overlapping, but subtly different patterns of expression that recapitulate endogenous *Nkx2–5* expression in the cardiac crescent and early heart tube (Searcy *et al.*, 1998; Lien *et al.*, 1999; Reecy *et al.*, 1999; Tanaka *et al.*, 1999). These enhancers also show activity in the pharynx, thyroid, stomach, and spleen, where *Nkx2–5* is expressed. Both of these enhancers contain essential binding sites for the GATA family of zinc-finger transcription factors (Searcy *et al.*, 1998; Lien *et al.*, 1999). However, binding of GATA factors alone cannot account for the specificity of *Nkx2–5* expression because GATA factors are expressed in many noncardiac tissues (Morrisey *et al.*, 1996, 1997).

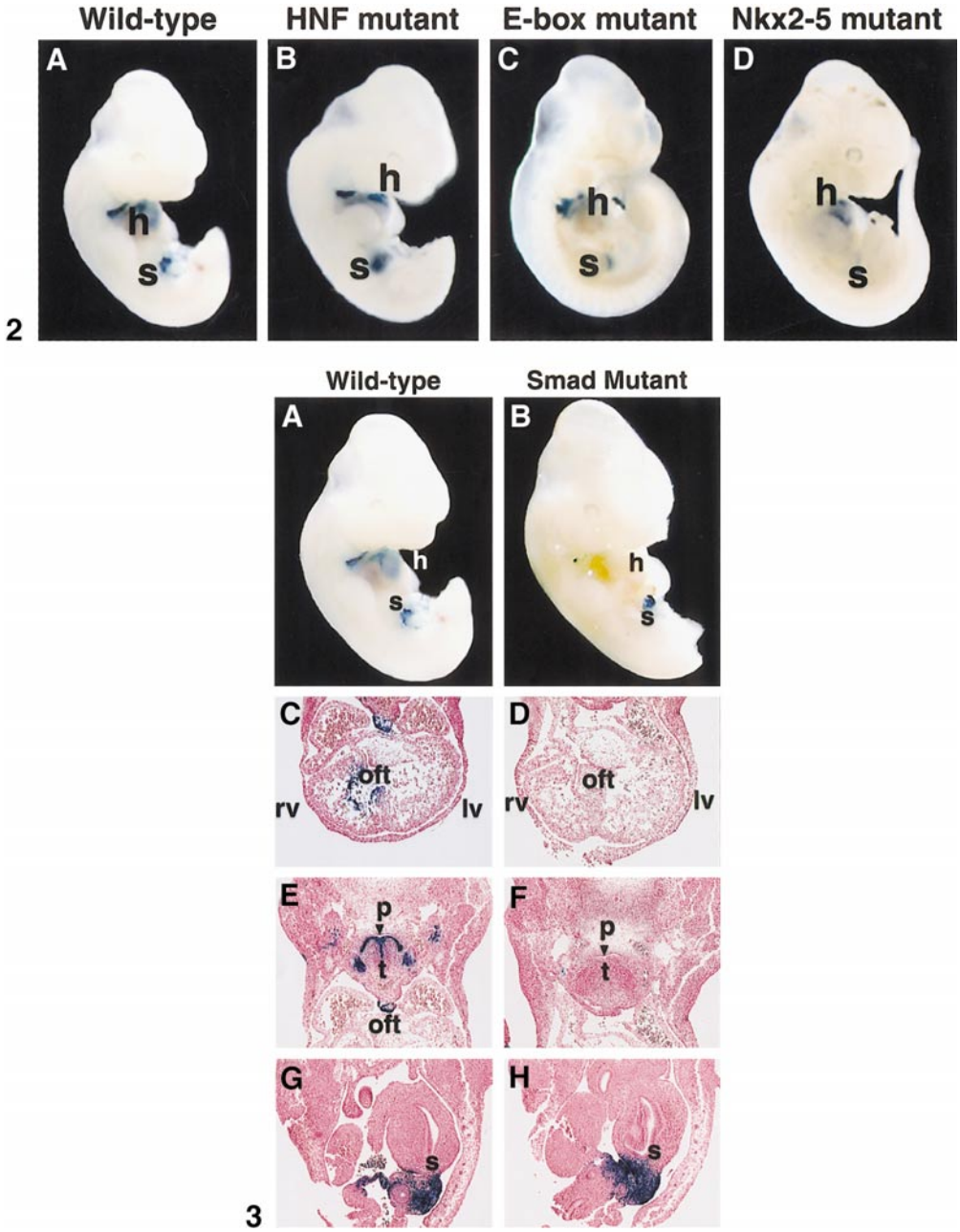

FIG. 2. Mutagenesis of HNF-3, Nkx2–5, and E-box consensus sequences in the *Nkx2–5* AR2 enhancer. F0 transgenic embryos harboring *lacZ* transgenes linked to (A) the wild-type *Nkx2–5* upstream region (−3050/+186) or the same region with mutations in the HNF-3 (B), E-box (C), or Nkx2–5 (D) sites were isolated at E11.5 and stained for lacZ expression. Sites are shown in Fig. 1B. The HNF-3 site at -3006 (CTCTGTTTGCTT) was mutated to CCCCGGGTGCTT. The E-box at −2841 (CATCTG) was mutated to CCCGGG. The Nkx2–5 site (AAAGTG) at −2943 was mutated to CCCGGG. h, heart; s, stomach.

FIG. 3. Expression of wild-type and Smad-mutant *Nkx2–5* enhancer at E11.5. F0 transgenic embryos harboring *lacZ* transgenes linked to the wild-type *Nkx2–5* upstream region (−3050/+186) (A, C, E, G) or the same region with a mutation in the Smad site at −2774 (B, D, F, H) were isolated at E11.5 and stained for lacZ expression. The posterior regions of the embryos in (A) and (B) were removed to allow visualization of the heart and stomach. (C–H) H&E sections of the embryos in (A) and (B) at different levels. Expression in the heart, pharynx, and thyroid is lost in the mutant, but expression in the stomach is retained. h, heart; lv, left ventricle; oft, outflow tract; p, pharynx; rv, right ventricle; s, stomach; t, thyroid.

Whether these enhancers can respond to early BMP signaling within the cardiac crescent or later has not been investigated.

To further understand the mechanisms that control *Nkx2–5* transcription in the developing heart, we analyzed the 5′-flanking regions of the mouse, chicken, and human *Nkx2–5* genes for sequences conserved between the three species, and we compared the *Nkx2–5* cardiac enhancers with the *tinman* tin-D enhancer. Here, we show that the AR2 enhancer of *Nkx2–5* is highly conserved in mouse, chicken, and human, and that this enhancer, like the tin-D enhancer, contains conserved Smad and Nkx2–5 binding sites. One of the Smad sites, which is adjacent to the two essential GATA sites (Searcy *et al.*, 1998), is essential for activity of the enhancer in the cardiac crescent and in later stages of heart development *in vivo*, but is not required for expression in the developing spleen or stomach. In contrast, the Nkx2–5 site is not required for enhancer activity at any developmental stage we analyzed. Our data suggest that *Nkx2–5* is a direct downstream target of Smad4, and also reveal differences between the transcriptional circuits that control *Nkx2–5* and *tinman* expression in the developing cardiac lineages of vertebrates and fruit flies.

MATERIALS AND METHODS

Cloning, Mapping, and Sequencing cNkx2–5 Genomic DNA

Four chicken *Nkx2–5* genomic clones were isolated from a chicken genomic library (Stratagene) by using a fragment of a chicken *Nkx2–5* cDNA as a probe. Positive clones were purified through secondary and tertiary screening and phage DNA was isolated. The genomic DNA of chicken *Nkx2–5* was subcloned into pBluescript, and 10 kb of 5′-flanking DNA was sequenced and compared with mouse sequences by using the FASTA programs provided in the GCG software package.

Plasmid Constructions

All mouse *Nkx2–5* nucleotide coordinates are related to the transcriptional start site. Transgenes were created by cloning DNA fragments into the *lacZ* expression vector pAUG-βGal. The enhancer fragment of the mouse *Nkx2–5* 5′-flanking region encompassing nucleotides −3050 to +186 relative to the transcriptional start site was PCR-amplified by using primer Y1 (5′-CCCAAGCTTGGGCTGCTCATCCATCAGCCAG) and Y2 (5′-CCCAAGCTTGGGCCAGGTGGGTAGCAGAGAG) and subcloned into pAUG-βGal. The enhancer-containing fragment −3050/+186/ΔSmad was generated by mutating the conserved Smad site at position −2774 to a *Sma*I site in the context of the −3050/+186 enhancer fragment. Similarly, the conserved HNF-3 binding site at −3006 (CTCTGTTTGCTT), the E-box at −2841 (CATCTG), and the Nkx2–5 binding site at −2943 (AAAGTG) were mutated to a *Sma*I site in the context of the same fragment.

Gel Mobility Shift Assay

Protein–DNA binding was assessed by gel mobility shift assays using glutathione *S*-transferase (GST) fusion proteins. The MH1 domain of Smad4 fused to GST was expressed from plasmid pGEX-Smad4 (MH1) in bacterial strain BL21-CodonPlus (Stratagene). Binding to the conserved Smad site located between nucleotides −2800 and −2750 was tested by using double-stranded DNA oligonucleotides with the sequence 5′-TAGGAAGATAAAGCTACGGATAACGCTGCCTGTCTCCAATGGACGTCTCCC; the conserved Smad site is underlined. Expression and purification of the fusion proteins were done according to the procedures of Guan and Dixon (1991), and gel shift assays were performed as described (Lien *et al.*, 1999). The mutant competitor oligonucleotide sequence is identical, except that the Smad site was mutated into a *Sma*I site. The Smad site was also tested for its binding to full-length GST-Smad1, Smad2, Smad3, and Smad4 (Cascade Bioscience).

Generation of Transgenic Mice, β-Galactosidase Staining, and Histology

DNA for pronuclear injection was gel-purified and eluted by using a QIAquick kit (Qiagen). Transgenic mice were generated and embryos were analyzed as described previously (Lien *et al.*, 1999).

RESULTS

Cross-Species Conservation of the Nkx2–5 Enhancer

The mouse *Nkx2–5* gene is controlled by a complex array of positive and negative regulatory regions located upstream, downstream, and within the gene (reviewed in Schwartz and Olson, 1999). The majority of *Nkx2–5* transcripts are encoded by two exons. However, a subset of transcripts is spliced to either of two alternative 5′ exons (Reecy *et al.*, 1999; Tanaka *et al.*, 1999). Exon 1a, the most 5′ alternative exon, is located approximately 4 kb upstream of the exon 1 (Fig. 1A). There are independent cardiac enhancers at −9437/−8922 and −3059/−2554 upstream of *Nkx2–5*, referred to as AR1 and AR2, respectively (Fig. 1A) (Searcy *et al.*, 1998; Lien *et al.*, 1999; Reecy *et al.*, 1999; Tanaka *et al.*, 1999; Schwartz and Olson, 1999). Throughout the text, we refer to DNA sequence positions relative to the major transcription initiation site at exon 1.

In an effort to identify possible evolutionarily conserved regulatory elements, we isolated chicken *Nkx2–5* genomic clones and compared the sequences of the chick and mouse 5′-flanking regions. Within 10 kb of 5′-flanking sequence, we identified a 470-bp segment with approximately 75% identity between the two species (Fig. 1B). This region of the mouse sequence, corresponding to the AR2 enhancer, directs *lacZ* reporter gene expression in the cardiac crescent and throughout the linear heart tube before becoming restricted to the outflow tract (Searcy *et al.*, 1998). This enhancer is also active in the pharynx, thyroid, stomach, and spleen (Searcy *et al.*, 1998; Lien *et al.*, 1999). We did not identify any other significantly conserved regions between chick and mouse sequences within the 10-kb upstream regions (Fig. 1A). Within the conserved enhancer region are two GATA sites that are essential for enhancer activity in the heart and spleen (Searcy *et al.*, 1998).

We also compared the mouse genomic sequence with the working draft sequence of human chromosome 5 (NT_023169.6; AC008412 (Sparrow *et al.*, 2000)) and thereby identified several stretches of homology between mouse and human sequences (Fig. 1A). Strikingly, the mouse and human sequences of the AR2 cardiac enhancer show a high degree of homology (Figs. 1A and 1B). In addition, we identified a 303-bp fragment within the AR1 enhancer and a short region between AR2 and the promoter that are also highly conserved between mouse and human (Fig. 1A; and data not shown). The sequence conservation of the two enhancers suggests that the mechanisms of *Nkx2–5* gene regulation are highly conserved between chicken, mouse, and human.

Mutagenesis of Transcription Factor Consensus Sites in the AR2 Enhancer

There are six conserved binding sites for GATA transcription factors in the AR2 enhancer (Fig. 1B), two of which have been shown to be essential for enhancer activity in heart and spleen (Searcy *et al.*, 1998). In an effort to identify other transcriptional activators of this enhancer, we focused on binding site consensus sequences that were conserved in the mouse, human, and chicken enhancers, including sites for forkhead transcription factor HNF-3, basic helix–loop–helix (bHLH) proteins, Nkx2–5, and Smads. As an initial test of the importance of these sites, we systematically mutated the binding site consensus sequences within the context of fragment −3050/+186, which contains the transcription start site and the AR2 enhancer, and analyzed the expression of a linked *lacZ* reporter gene in F0 transgenic mouse embryos at E11.5. At this stage, the −3050/+186 fragment directs reporter gene expression in the outflow tract of the heart, as well as in the pharynx, thyroid, and the distal portion of the stomach (which also eventually contributes to the spleen) (Searcy *et al.*, 1998; Lien *et al.*, 1999) (Figs. 2A, 3A, 3C, 3E, and 3G). A minimum of three independent F0 transgenic embryos was analyzed for each mutant construct.

A potential binding site for HNF-3, which is expressed in endoderm-derived tissues (Kaestner *et al.*, 1994), is present at −3006. An enhancer with a mutated HNF-3 site was fully active in the outflow tract heart, pharynx, thyroid, and stomach (Fig. 2B; and data not shown). This result indicated that this site was not required for AR2 enhancer activity in these tissues at E11.5.

Several bHLH transcription factors, which bind the E box consensus sequence (CANNTG), play important roles in cardiac development (Srivastava *et al.*, 1995; Nakagawa *et al.*, 1999). Therefore, the presence of a conserved E-box at −2841 of the AR2 enhancer suggested that the enhancer might be regulated by cardiac bHLH factors. However, the E box mutation did not abolish enhancer activity in the cardiac outflow tract or in other tissues at E11.5 (Fig. 2C; and data not shown).

It has been shown that *tinman* expression in the dorsal vessel of *Drosophila* requires binding of the Tinman protein itself to the tin-D enhancer, which has been proposed to fulfill a positive autoregulatory loop (Xu *et al.*, 1998). There are two binding sites for Tinman in the tin-D enhancer (Xu *et al.*, 1998). The presence of a highly conserved Nkx2–5 site at −2943 in the AR2 enhancer (Fig. 1B) suggested that, like Tinman, Nkx2–5 might regulate its own expression by binding to this sequence. However, an enhancer with this site mutated still directed *lacZ* expression in the outflow tract of the heart, pharynx, thyroid, and stomach (Fig. 2D). This result indicates that the Nkx2–5 site in the AR2 enhancer is not required for maintaining enhancer activity *in vivo*, and that unlike the tin-D enhancer of *tinman*, the AR2 enhancer is not directly regulated by Nkx2–5, at least at this stage of embryogenesis.

It should be pointed out that although the HNF-3, bHLH, and Nkx2–5 binding sites were not absolutely essential for enhancer activity at E11.5, we cannot rule out the possibility that they might be required for qualitative or quantitative effects on expression at other developmental stages. Indeed, the remarkable level of cross-species homology of these sites and surrounding sequences suggests their involvement in some aspects of *Nkx2–5* regulation.

A Highly Conserved Smad Site in the Nkx2–5 AR2 Enhancer Is Required for Enhancer Activity during Both Early and Late Cardiac Development

In addition to the transcription factor binding sites described above, there are multiple conserved GTCT (AGAC on the opposite strand) sequences at −3038, −3027, −2774, and −2758 in the AR2 enhancer that are highly conserved between chicken, mouse, and human sequences (Fig. 1B). These GTCT/AGAC sequences are potential binding sites (Zawel *et al.*, 1998; Johnson *et al.*, 1999) of Smad proteins, downstream effectors of TGFβ/BMP signaling (Figs. 1B and 1C). Among these potential Smad sites, the one at −2774 is adjacent to the two essential GATA sites (Searcy *et al.*, 1998). In addition, GTCT/AGAC boxes are often present as tandem repeats, and the Smad site (−2774) adjacent to the essential GATA sites consists of two tandem repeats of the sequence G(C/T)CTGTCT (Fig. 1B).

BMP has been shown to mediate induction of cardiogenesis and *Nkx2–5* expression in chick embryos (Schultheiss *et al.*, 1997; Andree *et al.*, 1998). However, it is unclear whether the regulation of *Nkx2*-5 by BMP signaling and its effectors is direct or indirect and whether *Nkx2–5* is regulated by BMP signaling in the mouse. The presence of conserved Smad binding sites in the chick and mouse enhancers suggests that mouse *Nkx2–5* might also be regulated by BMP signaling. To test whether the potential Smad site at −2774 was required for enhancer activity in the heart, we mutated this site in the context of fragment −3050/+186 and assayed for enhancer activity *in vivo*.

In contrast to the wild-type enhancer, which shows prominent expression in the outflow tract of the heart, pharynx, thyroid, and the distal part of the stomach at E11.5

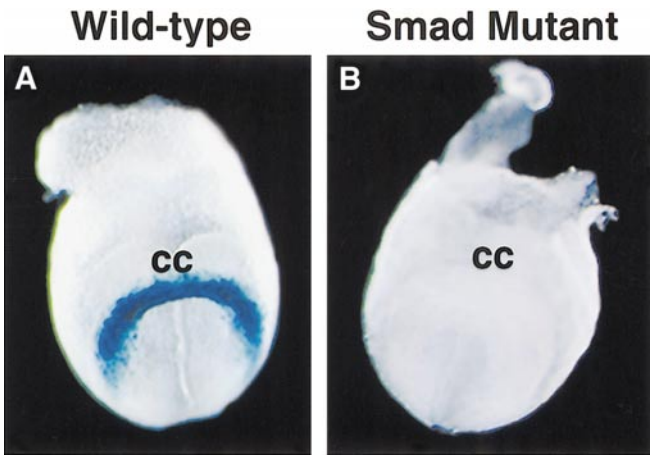

FIG. 4. Expression of wild-type and Smad-mutant *Nkx2–5* enhancer at E7.75. Embryos from transgenic mouse lines harboring *lacZ* transgenes linked to the wild-type *Nkx2–5* upstream region (−3050/+186) (A) or the same region with a mutation in the Smad site at −2774 (B) were isolated at E7.75 and stained for lacZ expression. All lacZ expression in the cardiac crescent (cc) is lost in the mutant.

(Searcy *et al.*, 1998; Lien *et al.*, 1999) (Figs. 3A, 3C, 3E, and 3G), no cardiac expression was detected in 10 F0 embryos harboring the enhancer with the mutated Smad site (Figs. 3B and 3D). That these embryos retained expression in stomach demonstrated that the transgene was functional and that this Smad binding site was not required for expression in these tissues (Fig. 3H). Interestingly, expression in the pharynx and thyroid was also abolished by the Smad site mutation (Fig. 3F; and data not shown).

To determine whether the Smad-mutant enhancer might be active at other stages of cardiac development, we generated three stable lines of transgenic mice harboring the mutant transgene. Like the F0 embryos, there was no cardiac expression of lacZ in these stable transgenic lines at E11.5, although expression in the stomach and spleen was retained (data not shown). Similarly, at E7.75, when the wild-type enhancer (−3050/+186 fragment) drives *lacZ* reporter gene expression in the cardiac crescent (Searcy *et al.*, 1998; Fig. 4A), there was no expression detected in the cardiac crescent in any embryos from the three stable lines harboring the transgene with the mutant Smad site (Fig. 4B). Thorough analysis of lacZ expression from these stable lines throughout embryogenesis and after birth also failed to reveal lacZ staining, in contrast to the wild-type enhancer, which is active at these stages (data not shown). These results demonstrate that the Smad site at −2774 is required for cardiac activity of the *Nkx2–5* AR2 enhancer from the earliest stage of cardiac specification *in vivo*. Interestingly, while BMP signaling has been implicated in activation of *Nkx2–5* in the cardiac crescent, the possibility that it might be required to maintain *Nkx2–5* expression at later stages has not been addressed. The finding that the Smad site is essential at both early and late developmental stages analyzed suggests a potential role for BMP signaling or other TGFβ signaling beyond the initial step in cardiogenic specification.

Binding of Smad4 to the Highly Conserved Smad Site in the Nkx2–5 AR2 Cardiac Enhancer

GTCT boxes can be bound by either Smad3 or Smad4 (Dennler *et al.*, 1998; Zawel *et al.*, 1998) or the Smad1/Smad4 complex (Johnson *et al.*, 1999). In light of the essential role of Medea, an ortholog of Smad4, as a direct regulator of the tin-D enhancer in *Drosophila* (Xu *et al.*, 1998), we tested whether mouse *Nkx2–5* was a direct target

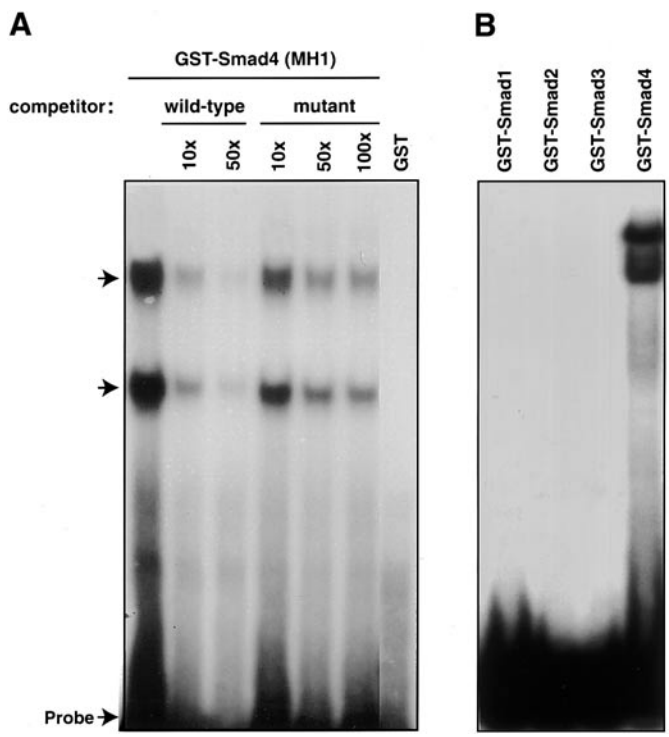

FIG. 5. Binding of Smad4 to the Smad site (−2774) in the *Nkx2–5* AR2 enhancer. (A) Binding of Smad4 (MH1) to the Smad site. A ^{32}P-labeled oligonucleotide probe encompassing the Smad site from the *Nkx2–5* AR2 enhancer was used in gel mobility shift assays with the MH1 domain of Smad4 expressed as a GST fusion protein, as described in Materials and Methods. The GST-Smad4 (MH1)-DNA complex was competed specifically with excess unlabeled probe, but was not effectively competed by a mutant probe. Fold-excess of competitor DNA compared with labeled probe is shown. The first lane on the left has no competitor. No DNA–protein complex was observed with GST alone. (B) Binding of full-length Smad4 to the Smad site. Full-length Smad1, Smad2, Smad3, and Smad4 were expressed as GST fusion proteins and gel mobility shift assays were performed as described above. Only GST-Smad4 bound to the Smad site.

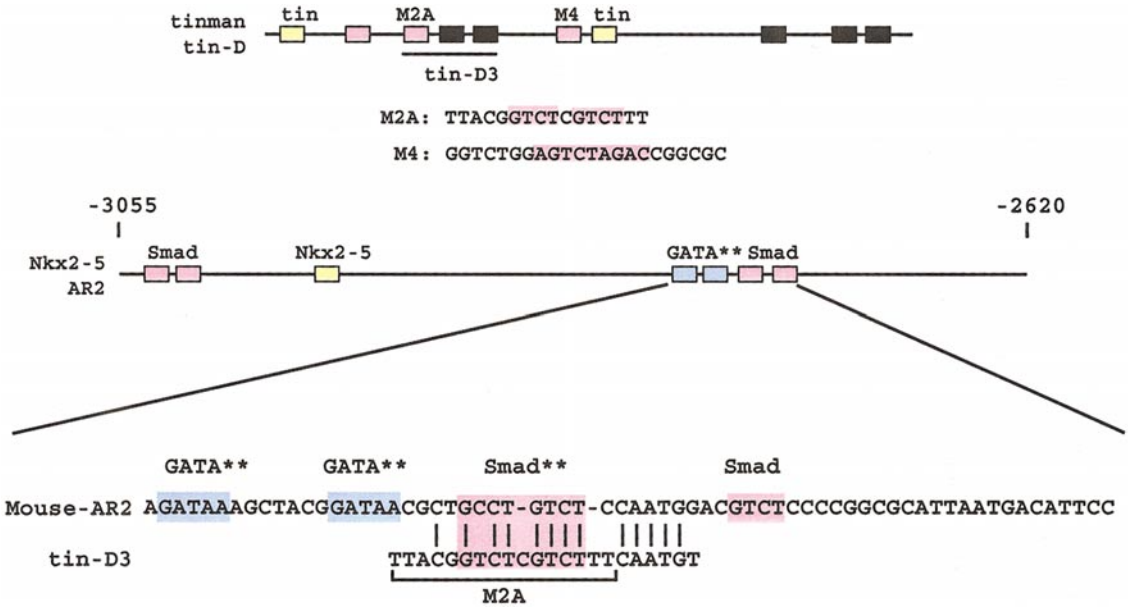

FIG. 6. Comparison of the *tinman* tin-D and *Nkx2–5* AR2 enhancers. The *tinman* and *Nkx2–5* enhancers are shown schematically. Tinman and Nkx2–5 binding sites are in yellow. Medea/Smad4 binding sites are shown in pink, and Mad binding sites are in black. The two essential GATA sites in the AR2 enhancer are marked with asterisks (**). The M2A and M4 sequences in the tin-D enhancer are Medea binding sites described in Xu *et al.* (1998) and Raftery and Sutherland (1999). The core GTCT sequences are indicated in pink beneath the tin-D schematic. Sequence homologies between the AR2 and tin-D enhancers in the region surrounding the essential Smad site in the AR2 enhancer are shown at the bottom. This region also shows homology to the core of the AR1 mouse enhancer (Lien *et al.*, 1999).

of Smad4. There are potential Smad3/4 binding sites at −3038, −3027, −2774, and −2758 in the AR2 enhancer that are highly conserved between chicken, mouse, and human sequences (Figs. 1B and 1C). Among these potential Smad sites, the one at −2774 is adjacent to the two essential GATA sites (Searcy *et al.*, 1998).

We tested whether Smad4 could bind to the conserved sequence at −2774, using GST-Smad fusion proteins and an oligonucleotide encompassing this putative Smad site in a gel mobility shift assay. As shown in Fig. 5A, a GST fusion protein containing the amino-terminal MH1 domain of Smad4, which can bind DNA in a ligand-independent manner (Zawel *et al.*, 1998), bound avidly to this DNA sequence. Binding was competed by unlabeled wild-type oligonucleotide, but was competed poorly by the same amount of oligonucleotide in which the Smad binding site was mutated. GST alone did not bind to the oligonucleotide. GST-Smad4 (MH1) bound to the oligonucleotide as two differently migrating complexes (Fig. 5A). Both complexes could be supershifted by Smad4 antibodies, suggesting that GST-Smad4 (MH1) may bind this sequence as both monomers and dimers (Fig. 5A).

To test whether full-length Smad4 and/or other Smad proteins can also bind to this element, we performed a gel mobility shift assay using GST fusion proteins of full-length Smad1, 2, 3, and 4. As shown in Fig. 5B, only GST-Smad4 was able to bind to the potential site, while Smad1, 2, and 3 had no binding activity. Therefore, we conclude that the GCCTGTCT sequence at −2774 in the AR2 enhancer is a cognate Smad4 binding site and that *Nkx2–5* is regulated directly by TGFβ/BMP signaling through Smad4.

DISCUSSION

The transcriptional regulation of mouse *Nkx2–5* is extremely complex, involving multiple positive and negative regulatory elements distributed in the 5′- and 3′-flanking regions of the gene (reviewed in Schwartz and Olson, 1999). Within the 23-kb genomic region examined in four studies (Searcy *et al.*, 1998; Lien *et al.*, 1999; Reecy *et al.*, 1999; Tanaka *et al.*, 1999), multiple enhancers that direct expression in heart, spleen, stomach, thyroid, and pharynx have been identified. In order to further delineate the cardiac enhancers, we compared the 5′-flanking regions of mouse, chick, and human *Nkx2–5*. Our results show that the AR2 enhancer, which is located 3 kb upstream of the transcription start site (Fig. 1A), is highly conserved between the three species. The AR1 enhancer, which is located 9 kb upstream (Fig. 1A), is also conserved between mouse and human (data not shown). However, we have not identified any homologous sequences to the AR1 enhancer within the 10-kb upstream region of chick *Nkx2–5*. Whether an AR1

homologous region is located outside this region of chicken genomic DNA remains to be determined. There is also a cardiac regulatory region located 887 bp upstream of alternative 5′ exon 1a (Reecy *et al.*, 1999), referred to as AR3 (Schwartz and Olson, 1999). We have not identified any sequence in chick or human genomic DNA with homology to the AR3 enhancer (Reecy *et al.*, 1999).

An important principle that has emerged from studies of other cell type-specific enhancers is that precise temporospatial control of transcription requires complex interactions among large numbers of transcription factors that bind overlapping and adjacent DNA sequences (reviewed in Davidson, 2000). While the studies reported here and elsewhere (Searcy *et al.*, 1998; Liberatore *et al.*, 2002) demonstrate the essential roles of the Smad and GATA sites for regulation of the AR2 enhancer, based on the extensive cross-species homology of this enhancer, it is likely that sequences throughout this region play important roles in fine-tuning the *Nkx2–5* expression pattern. Additional studies will be required to determine the identities of such factors and to understand the molecular basis for their interactions.

Comparison of the Regulation of Drosophila tinman and Mouse Nkx2–5 Expression

The organization of the *tinman* tin-D and vertebrate *Nkx2.5* enhancers is schematized in Fig. 6. There are four putative Smad4 binding sites, GTCT/AGAC, that are conserved in the AR2 enhancer (Fig. 1B). It has been shown that the *tinman* tin-D enhancer contains eight Mad binding sites, three of which can also be bound by Medea (Xu *et al.*, 1998). The consensus sequence of the Mad and Medea binding sites is the GC-rich sequence, CGCCGC (Xu *et al.*, 1998; Nguyen and Xu, 1998). However, for the sites that can also be bound by Medea, such as the M2 and M4 sites in the tin-D enhancer, there is an AGAC/GTCT sequence adjacent to the GC-rich sequence (Fig. 6; Xu *et al.*, 1998). This AGAC/GTCT sequence is identical to the vertebrate Smad4 binding site. Thus, it is likely that Medea actually binds to the AGAC/GTCT sequences instead of the GC-rich sequence.

Multiple Mad/Medea binding sites in the tin-D enhancer are required for dorsal mesoderm-specific activity of the enhancer (Xu *et al.*, 1998). In the AR2 enhancer, the Smad4 site at −2774 is required for enhancer activity in the cardiac crescent and later in heart development. These findings reveal an evolutionarily conserved role for Smad factors in the activation of cardiac NK-type homeobox genes, and support the notion that *Nkx2–5*, like *tinman*, is a direct target of Smad proteins.

Interestingly, when we compared the mouse AR1 and AR2 enhancers with the Dpp-responsive tin-D3 enhancer of *Drosophila* tinman, we found striking similarities among these enhancers. The essential Smad site at −2774 adjacent to the two essential GATA sites and the adjacent 3′-flanking sequences in the AR2 enhancer show high homology to the minimal Dpp response element in the tin-D enhancer (Xu *et al.*, 1998) (Fig. 6). In addition, the core of the mouse AR1 enhancer (Lien *et al.*, 1999) contains a region with high homology to the region surrounding the essential Smad site at −2774 in the AR2 enhancer. This putative Smad site is also close to the essential GATA site in the AR1 enhancer that we previously identified (Lien *et al.*, 1999). However, when we mutated this putative Smad site in the AR1 enhancer, enhancer activity was not abolished (E. Lien and E.N.O., unpublished), suggesting there might be other redundant Smad sites present in the AR1 enhancer.

Liberatore *et al.* (2002) have also examined the involvement of the Smad sites at the 5′ end of the AR2 enhancer. Consistent with our results, mutations of these binding sites eliminate activity of the enhancer in the cardiac crescent. However, these sites are not required for cardiac expression later in development and the mutant enhancer actually shows enhanced activity in the right ventricle, suggesting a negative role for Smad binding to these sites. Thus, it appears that the AR2 enhancer is a target for positive and negative regulation by Smad proteins at different stages of cardiac development. These divergent modes of regulation are likely to reflect differential associations of Smads with positive and negative cofactors that bind nearby sites in the enhancer.

Smads typically activate transcription in combination with other cofactors. Since BMPs are expressed in other regions of the embryo in addition to the cardiogenic region, the mechanism for BMP-dependent activation of *Nkx2–5* must be coupled to other cell-autonomous regulators expressed prior to *Nkx2–5*. Understanding how BMP signaling is interpreted in mesodermal cells by cardiogenic cofactors is likely to provide insights into the molecular basis for cardiac specification. In this regard, we have recently found that Smad4 interacts directly with GATA-4, providing a possible molecular basis for transcriptional synergy between these factors and for directly linking cardiac gene regulation with the BMP signaling pathway (E. Lien and E.N.O., unpublished).

Evolutionary Considerations

While the transcriptional regulation of *Nkx2–5* and *tinman* appear to be similar with respect to the dependence of the AR2 and tin-D enhancers on BMP signaling through Smad proteins, there are also fundamental differences in the regulation of these enhancers. In particular, the *tinman* tin-D enhancer is controlled through the combined actions of Medea and Tinman, whereas *Nkx2–5* does not seem to autoregulate its own expression through the Nkx2–5 binding site in the AR2 enhancer. On the contrary, it has been suggested that Nkx2–5 negatively regulates its own expression (Tanaka *et al.*, 1999), although we observed no evidence for enhanced expression of the enhancer with the Nkx2–5 binding site mutation, as might be predicted by such a model.

The differences in regulation of *tinman* and *Nkx2–5*

transcription reflect the differences in mesoderm specification and patterning of the vertebrate and arthropod body plans. *tinman* is expressed throughout the nascent mesoderm of *Drosophila* prior to its subdivision into different sublineages. Expression of *tinman* in the early mesoderm is mediated by binding of Twist to a separate enhancer (Yin *et al.*, 1997). Specification of the dorsal mesoderm occurs in response to Dpp signaling from the dorsal ectoderm. In contrast, *Nkx2–5* expression is initiated concomitant with cardiogenic specification in response to BMP signaling from the anterior endoderm. Thus, the mechanism for BMP-dependent activation must be coupled to other cell-autonomous regulators expressed prior to *Nkx2–5* itself. Understanding how BMP signaling is interpreted in mesodermal cells by cardiogenic cofactors is likely to provide insights into the molecular basis for cardiac specification.

ACKNOWLEDGMENTS

We thank Drs. J. Massagué and A. Hata for Smad plasmids, A. Lassar and K.-H. Lee for providing DNA sequence information, and K. Yutzey for sharing results prior to publication. We are grateful to Dr. M. Chao and A. Tizenor for assistance with graphics and members of the Olson laboratory for input and support. E.N.O. was supported by grants from the National Institutes of Health, the D. W. Reynolds Cardiovascular Clinical Research Center, the Robert A. Welch Foundation, and the W. G. McGowan Charitable Fund.

REFERENCES

Andree, B., Duprez, D., Vorbusch, B., Arnold, H. H., and Brand, T. (1998). BMP-2 induces ectopic expression of cardiac lineage markers and interferes with somite formation in chicken embryos. *Mech. Dev.* **70,** 119–131.

Attisano, L., and Wrana, J. L. (2000). Smads as transcriptional co-modulators. *Curr. Opin. Cell Biol.* **12,** 235–243.

Azpiazu, N., and Frasch, M. (1993). tinman and bagpipe: Two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. *Genes Dev.* **7,** 1325–1340.

Bodmer, R. (1993). The gene tinman is required for specification of the heart and visceral muscles in Drosophila. *Development* **118,** 719–729.

Bodmer, R. (1995). Heart development in Drosophila and its relationship to vertebrates. *Trends Cardiovasc. Med.* **5,** 21–28.

Bodmer, R., Jan, L. Y., and Jan, Y. N. (1990). A new homeobox-containing gene, msh-2, is transiently expressed early during mesoderm formation of Drosophila. *Development* **110,** 661–669.

Chen, C. Y., and Schwartz, R. J. (1995). Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. *J. Biol. Chem.* **270**, 15628–15633.

Davidson, E. (2000). "Genomic Regulatory Systems." Academic Press, New York.

Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J. M. (1998). Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. *EMBO J.* **17,** 3091–3100.

Fishman, M. C., and Olson, E. N. (1997). Parsing the heart: Genetic modules for organ assembly. *Cell* **91,** 153–156.

Franco, D., Kelly, R., Lamers, W. H., Buckingham, M., and Moorman, A. F. (1997). Regionalized transcriptional domains of myosin light chain 3f transgenes in the embryonic mouse heart: morphogenetic implications. *Dev. Biol.* **188,** 17–33.

Frasch, M. (1995). Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early Drosophila embryo. *Nature* **374,** 464–467.

Guan, K. L., and Dixon, J. E. (1991). Eukaryotic proteins expressed in *Escherichia coli*: An improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* **192,** 262–267.

Harvey, R. P. (1996). NK-2 homeobox genes and heart development. *Dev. Biol.* **178,** 203–216.

Hata, A., Seoane, J., Lagna, G., Montalvo, E., Hemmati-Brivanlou, A., and Massagué, J. (2000). OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP- Smad and Olf signaling pathways. *Cell* **100,** 229–240.

Johnson, K., Kirkpatrick, H., Comer, A., Hoffmann, F. M., and Laughon, A. (1999). Interaction of Smad complexes with tripartite DNA-binding sites. *J. Biol. Chem.* **274,** 20709–20716.

Jonk, L. J., Itoh, S., Heldin, C. H., ten Dijke, P., and Kruijer, W. (1998). Identification and functional characterization of a Smad binding element (SBE) in the JunB promoter that acts as a transforming growth factor-beta, activin, and bone morphogenetic protein-inducible enhancer. *J. Biol. Chem.* **273,** 21145–21152.

Kaestner, N. H., Hiemisch, H., Luckow, B., and Schutz, G. (1994). The HNF-3 gene family of transcription factors in mice: Gene structure, cDNA sequence, and mRNA distribution. *Genomics* **20,** 377–385.

Komuro, I., and Izumo, S. (1993). Csx: A murine homeobox-containing gene specifically expressed in the developing heart. *Proc. Natl. Acad. Sci. USA* **90,** 8145–8149.

Kuisk, I. R., Li, H., Tran, D., and Capetanaki, Y. (1996). A single MEF2 site governs desmin transcription in both heart and skeletal muscle during mouse embryogenesis. *Dev. Biol.* **174,** 1–13.

Li, L., Miano, J. M., Mercer, B., and Olson, E. N. (1996). Expression of the SM22alpha promoter in transgenic mice provides evidence for distinct transcriptional regulatory programs in vascular and visceral smooth muscle cells. *J. Cell Biol.* **132,** 849–859.

Liberatore, C. M., Searcy-Schrick, R. D., Vincent, E. B., and Yutzey, K. E. (2002). Nkx-2.5 induction in mice is mediated by a Smad consensus regulatory region. *Dev. Biol.*, this issue.

Lien, C. L., Wu, C., Mercer, B., Webb, R., Richardson, J. A., and Olson, E. N. (1999). Control of early cardiac-specific transcription of Nkx2–5 by a GATA- dependent enhancer. *Development* **126,** 75–84.

Lints, T. J., Parsons, L. M., Hartley, L., Lyons, I., and Harvey, R. P. (1993). Nkx-2.5: A novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* **119,** 969.

Massagué, J., and Chen, Y. G. (2000). Controlling TGF-beta signaling. *Genes Dev.* **14,** 627–644.

Morrisey, E. E., Ip, H. S., Lu, M. M., and Parmacek, M. S. (1996). GATA-6: A zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.* **177,** 309–322.

Morrisey, E. E., Ip, H. S., Tang, Z., Lu, M. M., and Parmacek, M. S. (1997). GATA-5: A transcriptional activator expressed in a novel

temporally and spatially-restricted pattern during embryonic development. *Dev. Biol.* **183,** 21–36.

Nakagawa, O., Nakagawa, M., Richardson, J. A., Olson, E. N., and Srivasta, D. (1999). HRT1, HRT2, and HRT3: A new subclass of bHLH transcription factors marking specific cardiac, somitic, and pharyngeal arch segments. *Dev. Biol.* **216,** 72–84.

Nascone, N., and Mercola, M. (1996). Endoderm and cardiogenesis: New insights. *Trends Cardiovasc. Med.* **6,** 211–216.

Nguyen, H. T., and Xu, X. (1998). *Drosophila mef2* expression during mesoderm development is controlled by a complex array of cis-acting regulatory modules. *Dev. Biol.* **204,** 550–556.

O'Brien, T. X., Lee, K. J., and Chien, K. R. (1993). Positional specification of ventricular myosin light chain 2 expression in the primitive murine heart tube. *Proc. Natl. Acad. Sci. USA* **90,** 5157–5161.

Park, M., Lewis, C., Turbay, D., Chung, A., Chen, J. N., Evans, S., Breitbart, R. E., Fishman, M. C., Izumo, S., and Bodmer, R. (1998). Differential rescue of visceral and cardiac defects in Drosophila by vertebrate tinman-related genes. *Proc. Natl. Acad. Sci. USA* **95,** 9366–9371.

Raftery, L. A., and Sutherland, D. J. (1999). TGF-beta family signal transduction in Drosophila development: From Mad to Smads. *Dev. Biol.* **210,** 251–268.

Ranganayakulu, G., Elliott, D. A., Harvey, R. P., and Olson, E. N. (1998). Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice. *Development* **125,** 3037–3048.

Reecy, J. M., Li, X., Yamada, M., DeMayo, F. J., Newman, C. S., Harvey, R. P., and Schwartz, R. J. (1999). Identification of upstream regulatory regions in the heart-expressed homeobox gene Nkx2–5. *Development* **126,** 839–849.

Schultheiss, T. M., Burch, J. B., and Lassar, A. B. (1997). A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev.* **11,** 451–462.

Schultheiss, T. M., Xydas, S., and Lassar, A. B. (1995). Induction of avian cardiac myogenesis by anterior endoderm. *Development* **121,** 4203–4214.

Schwartz, R. J., and Olson, E. N. (1999). Building the heart piece by piece: Modularity of cis-elements regulating Nkx2–5 transcription. *Development* **126,** 4187–4192.

Searcy, R. D., Vincent, E. B., Liberatore, C. M., and Yutzey, K. E. (1998). A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. *Development* **125,** 4461–4470.

Shi, Y., Katsev, S., Cai, C, and Evans, S. (2000). BMP signaling is required for heart formation in vertebrates. *Dev. Biol.* **224,** 226–237.

Sparrow, D. B., Cai, C., Kotecha, S., Latinkic, B., Cooper, B., Towers, N., Evans, S. M., and Mohun, T. J. (2000). Regulation of the tinman homologues in Xenopus embryos. *Dev. Biol.* **227,** 65–79.

Srivastava, D., Cserjesi, P., and Olson, E. N. (1995). A subclass of bHLH proteins required for cardiac morphogenesis. *Science* **270,** 1995–1999.

Tanaka, M., Wechsler, S. B., Lee, I. W., Yamasaki, N., Lawitts, J. A., and Izumo, S. (1999). Complex modular cis-acting elements regulate expression of the cardiac specifying homeobox gene Csx/Nkx2.5. *Development* **126,** 1439–1450.

Xu, X., Yin, Z., Hudson, J. B., Ferguson, E. L., and Frasch, M. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. *Genes Dev.* **12,** 2354–2370.

Yin, Z., Xu, X. L., and Frasch, M. (1997). Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* **124,** 4971–4982.

Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B., and Kern, S. E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell.* **1**, 611–617.

Zhang, Y., Feng, X. H., and Derynck, R. (1998). Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature* **394,** 909–913.

Received for publication January 4, 2002
Accepted January 21, 2002
Published online March 14, 2002